DOCUMENT RESUME

ED 166 248 95 TM 008 282

TITLE	Proceedings of the National Conference on Minimum Competency Testing.
INSTITUTION	Northwest Regional Education Lab., Portland, Oreg. Clearinghouse for Applied Performance Testing.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Oct 78
NOTE	92p.; Papers presented at the National Conference on Minimum Competency Testing (Toronto, Ontario, Canada, March 29, 1978); Not available in hard copy due to author's restrictions
AVAILABLE FROM	Clearinghouse for Applied Performance Testing, Northwest Regional Educational Laboratory, 710 S.W. Second Avenue, Portland, Oregon 97204 ($3.00)
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; *Academic Standards; *Basic Skills; Conference Reports; *Cutting Scores; Elementary Secondary Education; Equated Scores; Legal Problems; Mastery Tests; *Minimum Competency Testing; Political Issues; State Programs; *Student Certification; Test Bias; *Testing Problems; Test Interpretation
IDENTIFIERS	California

ABSTRACT

Seven prepared papers, with questions from the audience and answers by participants make up the proceedings of this conference on minimum competency testing. A number of conceptual and technical problems were raised, as well as questions about the effects of minimum competency testing on the curriculum and the psychological effects of the program on students, teachers, and others. The moral and legal problems concerned with phase-in periods, the match between the test and the curriculum, the possibility of racial discrimination, and the characteristics of a model program were discussed. Techniques for implementing cut-off decisions were described and the place of human judgment was debated. A suggestion was made that scaled certification was preferable to categorical pass-fail decisions. Comments were made on a series of possible positive and negative effects of minimum competency testing. The California state testing program was described, and questions were raised about the process and validity of setting standards. The speakers included Decker F. Walker, Merle Steven McClung, Carol A. Dwyer, Loretta A. Shepard, Ellis B. Page, Ruth S. Nickse, and Dale Carlson. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

- ED166248

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
NWREL
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Proceedings of the National Conference on Minimum Competency Testing

TM008 282

Clearinghouse for Applied Performance Testing
Northwest Regional Educational Laboratory
Portland, Oregon

2

Proceedings of the National Conference on Minimum Competency Testing

Clearinghouse for Applied Performance Testing
Northwest Regional Educational Laboratory
Portland, Oregon

October 1978

Published by the Northwest Regional Educational Laboratory, a private nonprofit corporation. The project presented herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Copyright © 1978 by the Northwest Regional Educational Laboratory. All Rights Reserved.

TABLE OF CONTENTS

	Page
PREFACE	v
PART ONE: INVITED PRESENTATIONS	
The Impact of Minimum Competency Testing on School Curricula, Teaching, and Students' Learning	
Decker F. Walker	3
Are Competency Testing Programs Fair? Legal?	
Merle Steven McClung.	13
A Debate on the Proposition: Adequate Measurement Technology Exists to Implement Fair, Equitable, and Useful Minimum Competency Testing Programs	
Carol A. Dwyer Loretta A. Shepard.	27
PART TWO: PANEL REACTIONS	
Escaping the Minimum Competency Dilemma: Scaled Certification	
Ellis B. Page	51
Comments on Minimum Competency Testing	
Ruth S. Nickse.	61
Dale Carlson.	69
PART THREE: QUESTIONS AND ANSWERS.	79

PREFACE

On March 29, 1978, the Clearinghouse for Applied Performance Testing (CAPT) brought together several prominent educators to present their views on minimum competency testing. This conference, held in Toronto during the AERA/NCME Annual Meetings, attracted nearly 400 educators from across the country.

In planning the conference, CAPT staff decided to organize it around three major issues associated with minimum competency testing: measurement issues; legal considerations; and implications for curriculum, instruction, and student learning. Speakers were selected on the basis of their ability to respond to these issues.

Dr. Decker F. Walker, Associate Professor of Education at Stanford University, discussed "The Impact of Minimum Competency Testing on School Curricula, Teaching and Students' Learning." Dr. Walker recently served as Vice President of AERA Division B (Curriculum and Objectives), and has presented his views on minimum competency testing in the November 1977 issue of Educational Leadership.

Merle Steven McClung, part-time Staff Attorney at the Center for Law and Education and part-time education law consultant, spoke on the topic "The Legal Issues Associated with Minimum Competency Testing." Mr. McClung has recently served as a legal advisor to states designing and implementing minimum competency testing programs. He has also published articles on minimum competency testing in Phi Delta Kappan and Clearinghouse Review. At the request of Mr. McClung, his article

"Are Competency Testing Programs Fair? Legal?" which appeared in the February 1978 Phi Delta Kappan, is reprinted in the Conference Proceedings in lieu of his oral presentation.

To examine the measurement issues, CAPT organized a debate on the proposition, "Adequate Measurement Technology Exists to Implement Fair, Equitable and Useful Minimum Competency Testing Programs." The affirmative stance was taken by Dr. Carol Dwyer, Director of the Basic Skills Assessment at Educational Testing Service. In that role, Dr. Dwyer supervises a basic skills testing system compatible with many states' and districts' minimum competency testing programs.

Dr. Loretta Shepard, Assistant Professor of Education at the University of Colorado, and editor of the Journal of Educational Measurement, took the negative stand. Dr. Shepard, who has published articles offering educators advice on setting standards, recently edited an issue of JEM devoted to minimum competency testing.

To stimulate further discussion, CAPT invited a panel of three individuals to react to the presentations and provide their own insights. Panel members included Dr. Ellis B. Page, Professor of Educational Psychology at the University of Connecticut and President-Elect of AERA; Dr. Ruth Nickse, Director of Nickse Associates; and Dr. Dale Carlson, Assistant Chief, Office of Program Evaluation and Research, California Department of Education.

We are deeply indebted to the Conference's participants, not only for the excellence of their presentations but also for their help in preparing the Proceedings. By examining the complexities of minimum competency testing from many points of view, they have advanced our understanding of this educational phenomenon.

PART ONE

INVITED PRESENTATIONS

The Impact of Minimum Competency Testing on School Curricula, Teaching, and Students' Learning

Decker F. Walker
Stanford University

Minimum competency testing is shaping up as the next major reform in American education. Legislatures in several large states have already passed laws establishing minimum competency requirements for promotion and graduation. Sometimes, as in my home state of California, these laws require all students to pass tests in reading, writing, arithmetic or other academic skills. Sometimes, as in my neighboring state of Oregon, the law requires that competencies be more practical--balancing a checkbook, writing a letter, applying for a job, and the like. Even in states that do not mandate competency testing statewide, local districts may adopt it. The main support for the movement toward competency-based education seems to come from laymen rather than the professional educators. But many professionals are also concerned about youngsters who leave school unable to read, write, compute, or meet the other minimum-skill demands of daily life in our complex society.

For my part, I'm worried about minimum competency testing in education. Before I can support it, I'd like to have satisfying answers to a number of questions that seem to me to be very important: How do we determine minimum competency? Are the three R's sufficient? What about practical skills, American history, civics and government, career

entry skills? What level should be set as minimum in any or all of these areas? Should students be able to spell 90% correctly, 100% correctly, 75% correctly, or what? Should we insist that they be able to read at the level of TV ads and highway signs? At the level of the daily newspaper? Or read and understand the Constitution? Who's to make these decisions? Shall we have a vote of the people? Shall we have experts' determinations? And if so, which experts? Shall local teachers decide? Or state bureaucracies? Or federal agencies? How are we to avoid the rigidities of a national system of minimums, and at the same time avoid the inequities and the chaos that would come from thousands of conflicting standards in different states and localities? Can we afford to develop reliable, valid tests that would correspond to every district's individual preferences for minimum standards? Or will economic pressures and public demands for equity not eventually force us to a nationwide set of standards? How are we to cope fairly with all the special circumstances that would threaten the validity of a minimum competency test--including such well-known factors as test anxiety, bilingualism, or learning disorders of various kinds?

It's not surprising to me that laypeople would overlook or discount or be entirely unaware of such conceptual and technical problems. Equally unsurprising is that professionals would bring these problems to the forefront of their attention--because it is they who must implement minimum competency testing, who will somehow have to deal with the problems raised by each of those questions and many others we haven't thought about. From the public's perspective, however, the problem is a simple one. We must simply see to it that all children master basic skills, and the obvious remedy to a non-professional layperson looking

at the situation is simply to insist by means of a test of some sort that every student has achieved those skills.

If we somehow manage to cope with the conceptual and technical problems indicated in the questions I've just asked and manage to implement a fair and effective system of competency-based education, will the overall impact of that system on the school, on the children, and on society in the short term and in the long term be positive and humane? I can't be quite certain that it would.

Will competency tests, for example, have the effect of labeling children, as young as age 13 or so, as incompetent, not up to minimum? And perhaps segregating these children from their peers to give them special help that they need? Will the failures of the system be primarily the children of the poor, of the minorities, of the disadvantaged families? If they are, will we again be blaming the victims, rather than attacking the cause of the problem? What will happen to those who fail to meet our standards? Will they be denied a diploma? And if so, what will that do to their employability, which surely in low achieving youngsters is already a difficulty in their lives? Now we've added to that difficulty the additional obstacle of no credential, no diploma. Have we not doomed these individuals to virtual unemployability? And if so, what kind of critical problems are we creating in doing so? Alternatively, will we retain these failures with the younger students in the next year's class, and the next year's, and the next year's? We know of the great disruptions caused by older students retained in classes from experience around the turn of the century in American schools. Will it then be possible to sustain a humane school environment while operating a minimum competency testing program? I am not quite sure.

Because of my fears, I am not enlisting in the movement for competency testing. On the other hand, I'm not ready to denounce it either. I cannot dismiss the equally plausible fears of the movement's supporters, primarily fears that each year thousands of students advance through our schools without the basic equipment they need for making their way in the world. Those of us who work in public education serve the public. It is irresponsible for us to ignore and dismiss the public's rightful, serious and longstanding concern over the quality of basic skills instruction that children receive. If we only knew what the impact of competency testing would actually be, and we knew the pros and cons of it, then perhaps we could make up our minds on whether we would support minimum competency testing or not, in what way and in what form.

That is ostensibly the title of my remarks before you this morning--the impact of competency testing on curriculum and students. Unfortunately, neither I, nor as far as I am able to tell, anyone else, can predict with any degree of certainty what the consequences would be of implementing minimum competency testing. Even though we have some historical examples and some current examples of operating systems, it is still not clear what the consequences would be nationwide.

It is my belief that in principle we won't be able to predict what the consequences would be. I think the metaphor implicit in terms like "impact" or "effect" is basically one of mechanical cause and effect. You do this and the consequences then ensue. Just as when you release a bowling ball--as soon as you release it, mathematically we can predict the result at the end of the alley.

I don't think that sort of thing, or anything really close to it, happens in a situation like the one we are talking about. I think the situation is somewhat more like a sociodrama, in which you have certain

actors who have parts to play. In this case they have the parts to start with and you confront them with a situation--a problem that they have to face. Now, they in their interaction work out the play as they go along. In this case the main actors would be the students. All students would be affected, because in most instances, all their students would be tested. Most administrators would be affected by minimum competency testing in one form or another. And of course the lay public would also be affected by it.

Those would be the main actors in this little sociodrama, and the problem we confront them with is some scheme of minimum competency testing. I submit to you that if you wanted to find out what the possible effects of minimum competency testing on the curriculum or students would be, probably the best way to do it would be to actually enact this little sociodrama. You could get some real students, some real teachers, some real parents and community members, and some real administrators and give them a minimum competency testing plan. You could give them a day to study it among themselves to figure out what's in it for them, and what the dangers are in the plan. Then you could get them on stage and ask them to role play some situations for us--ask them "what would you do? How would you feel?"

You can see how it might go--I can give you a few thoughts on the possibilities. The students might be pleased by the prospect of getting special help, especially the youngsters who are very anxious about whether they want to get through school, whether they want to learn to read, and so on. It might very well be that those youngsters will express pleasure and satisfaction in the plan at the prospect that they

would get more help, not be passed over and not be left as failures, but be remediated in some way.

On the other hand, the students might be highly anxious about the test. To illustrate the difficulties of high test anxiety (especially when the student may indeed be disgraced by being put into a special category of students and potentially separated from his peers if he performs below standard). Consider the following analogy. Imagine a board lying on the floor, say six inches wide and about six feet long, and you're asked to walk the length of it--that's your competency test. I doubt any of us would have any problems doing it. But now raise the same board about 100 feet and don't put any net down below. The anxiety of possibly suffering the severe consequences would make it very difficult for almost any of us to walk across that plank. I don't know how many youngsters will perceive it as a six inch wide plank lying on the ground. But the danger is there, and the students would play all this out for us.

There's the problem of labeling some students as failures, potentially segregating them. There's already a tendency, as all of us know, for students to speak to one another in derogatory terms--"dummies" and so on. Now we officially say "These <u>are</u> the dummies. They can't pass the minimum competency test." Thus we can't imagine the students doing anything other than increasing their own expressions of derision and veiled hostility and anxiety toward the students who are labeled as "dummies".

We can talk about the time that would be involved in testing for the students. Some of these students, perhaps the brighter ones, might very well say "Why should I spend two hours in the 9th grade taking

these dumb tests, when I know I can do them right now, when I know that it's just taking time away from things I can do that would be more fun and more educative for me? Why should I waste my time?"

Perhaps the most serious problem might be the denial of a diploma. I can imagine minority students, let's say, who would role-play having been denied a diploma. They passed all their courses, they sat through 12 years of school, and they completed assignments to at least a passing level for all their teachers. The time comes for graduation, and they do not pass the test and the diploma is denied. You can easily see them going to court, and saying, "Look--prove to me that passing this test that you've established, this minimum competency test, is really minimally required for success in the society. I don't think it is. I can get along just fine without it. I passed all my courses without it and I can get a job without it (if you'd give me a diploma). I don't think it's as necessary as you say." Now I certainly don't want to venture into predicting what the court would decide in such a case, but there would certainly be such cases.

What kinds of things might teachers say in this role playing session? Well, they might be happy to get special help for their low achieving youngsters. However, they might be a little bit anxious about the possibility of having extra duties with respect to the low achieving youngsters. They would lose time from the classroom. Teachers are already complaining, individually and in their organizations, about the amount of time given over to testing and the incursions that makes on classroom time. Here would be additional testing time. They would be concerned about how they could manage the youngsters' anxiety, because it disrupts their classroom. They would also be concerned about whether

they would be blamed for their students' failure. If five of my students fail to pass the minimum competency test this year, and that's more than in anybody else's class, won't that look very unfavorable to me as a teacher? Won't that mean, therefore, that I would strive to get the ablest students and the fewest of the "dummies" that I could, because I don't want them to make me look bad next year? That might create some tensions and internal problems as teachers try to jockey for the best prospects in their classes.

There may be some positive effects for teachers in opportunity for professional advancement. Any time you have a new program, a teacher who is closely identified with it, who gets to work with it, and is innovative with it may indeed find it an avenue for promotion within the school system. They can aim for "Director of Minimum Competency Testing" or something of that sort within the organization.

Well, I'm going to stop doing this. I think you can all follow out these scenarios just as well as I can. It would be much better if we could have the people up here and actually work them out. What actually would come to pass when we have a minimum competency test, of course, depends a great deal on the specifics, and no one can predict those. It depends on exactly what kind of minimum competency testing plan you're talking about--they really differ quite a bit. Are we talking about tests that are strictly academic, strictly practical, a mixture of academic and practical? Are we talking entirely about paper and pencil tests, or are we also going to consider other kinds of performance tests? Are we talking about relatively easy minimums that will be a threat to only the lowest 10 percent of the population? Or are we talking about much higher minimums that may be a threat to

perhaps 30 or 50 percent of the school population? And how are the minimums established? If the minimums have a great deal of support from the public and they seem to be valid I would imagine one kind of consequence. On the other hand, if they seem to be arbitrary, if they don't seem to have any rationale, and if many groups in the population don't accept the standards, I'm sure you can imagine another set of consequences.

So I'm afraid I've gone about as far as I can go toward helping you to think about the impact of minimum competency testing. I don't think that we can tell just yet what the impact is likely to be. There are a range of possibilities. I hope that the positive ones come to pass. For me, the most positive one would be that minimum competency testing turns out to be a rather specialized tool for identifying those low achieving youngsters in school who are really having problems with functional literacy and with survival in day to day tasks. Those youngsters are identified and given some special help, with minimal impact on the rest of the school. From any point of view, mass testing of every youngster in the school just really doesn't make any sense. It doesn't make any sense when it comes to the economic use of tests and test information. It seems to me that the goals of minimum competency testing can best be achieved by concentrating our test energies and our remedial energies on the youngsters who are most at risk. It's as though we had a mass screening of the population for appendicitis all the time, when we really only need to deal with the people who have the symptoms of appendicitis. It's unrealistic.

I would hope that some kind of positive scenario would eventuate which would identify these youngsters, give them special help, and

reduce the number of them who leave school functionally illiterate and unable to cope. I guess if I had to make an ultimate judgement, I would say that in my estimation the probability of that positive result would not exceed 25 percent. There are so many pitfalls, so many difficulties, so many things that would have to be controlled and handled--including the public's fervor for minimum competencies. In many communities such as my own home town of Palo Alto, this fervor is not for minimums, as I would understand them, but for the highest possible minimums. We in Palo Alto don't want just any old minimum--we want a minimum that really shows that our students have a great deal of knowledge when they leave school. I suspect that this kind of competitive ethic of "our minimums are better than your minimums," is likely to come about.

These, and the other problems already mentioned, and many others that I am not able to foresee, constitute grave dangers in which I fear we may stumble.

Are Competency Testing Programs Fair? Legal?

Merle Steven McClung
Center for Law and Education

This article appeared in the Phi Delta Kappan, February 1978, and is reprinted here by permission of the author and Phi Delta Kappan.

Responding to the public's demand for accountability in teaching and learning basic skills, legislators and educators in many states have unfairly shifted the burden of poor performance from the schools to the students. An increasing number of states and school districts are requiring students to pass a minimal proficiency or competency test as a prerequisite to a high school diploma. Given the number of studies showing substantial numbers of functionally illiterate students receiving high school diplomas, the requirement on its face seems fair enough. Many of the programs incorporating this requirement, however, are designed and/or implemented in an inequitable manner, and some may be illegal. This article will discuss these points in more detail under the following headings: inadequate phase-in periods, inadequate match between test and instruction, racial discrimination, and model program provisions.

Inadequate Phase-In Periods

Many competency programs are being imposed upon students late in their secondary education with little prior notice. Imposition one year before graduation means that a student will have spent his first 10 or 11 years in the school system without notice or knowledge that a passing competency test would be a condition for acquiring the diploma. In

fact the school district would have explicitly approved his progress by promoting him each year, even if he did not have basic competencies. It is likely that many if not most of those students failing the test would have studied differently in earlier years had they been given such notice--and teachers might have taught differently as well. The competency test is designed to assure that minimal competency is acquired after 12 years of schooling, but students in this situation would not have received notice until their tenth or eleventh year of schooling.

Traditional notions of due process should require adequate prior notice of any rule that could cause irreparable harm to a person's educational or occupational prospects. Whatever notice is considered fair in this situation (first grade? fourth grade?), notice after most of one's educational program is already completed clearly seems inadequate.

Inadequate Test/Instruction Match

Most persons would agree that fairness requires a school's curriculum and instruction to be matched in some way with whatever is later measured by the test. In other words, the test would be unfair if it measured what the school never taught. This concept should be considered in terms of both curricular validity and instructional validity.

Curricular validity is a measure of how well test items represent the objectives of the curriculum. An analysis of curricular validity would require comparison of the test objectives with the school's course objectives. For example, if the curriculum was not designed to teach functional competency, it would be unfair to deny individual

students their diploma because they did not learn to be functionally competent. In this situation, failure on the competency test should reflect on the schools, which are not offering an appropriate curriculum.

A competency test should also have what may be called instructional validity. Even if the curricular objectives of the school correspond with the competency test objectives, there must be some measure of whether or not the school district's stated objectives were translated into topics actually taught in the district's classrooms. While a measure of curricular validity is a measure of the theoretical validity of the competency test as an instrument to assess the success of students, instructional validity is an actual measure of whether the schools are providing students with instruction in the knowledge and skills measured by the test. Instructional validity obviously does not require prior exposure of the student to the exact questions asked on the test, but it does require actual exposure of students to the kind of knowledge and skills that would enable a student to answer the test questions. This will present difficult proof problems in some cases; in others, it will be relatively easy to show that the test is measuring what the school never taught.

It is important to note that content validity, as defined by the American Psychological Association,[1] does not insure either curricular or instructional validity. The concepts are related but distinguishable. Content validity is a measure of how well test items represent the performance domain that the test purports to measure (for example, adult performance skills), but it is not necessarily a measure of how well the test items and performance domain represent either a particular

school's curricular objectives or instruction received. Instructional validity should be the central concern, because content and curricular validity mean very little in this context if the test items are not representative of instruction actually received by the student.

Test developers will be reluctant to evaluate the curricular or instructional validity of their tests, but they have some kind of professional responsibility (beyond the usual disclaimers) to assure that their tests do not measure skills and knowledge that were never taught in school, particularly when they design a test instrument for an individual user. A promising step in the direction of professional responsibility was recently taken by the Educational Testing Service (ETS) when it announced that it would not report National Teacher Examination (NTE) scores to the South Carolina Department of Education until it had assurance against probable misuse.[2] The ultimate responsibility for the curricular and instructional validity of the test, however, should be with the school or school system that uses the test. A school or school system that cannot assure the curricular and instructional validity of its competency tests should not use them as a basis for denying promotion or a diploma to any of its students.

Questions of curricular and instructional validity are relevant, given either of two general purposes of competency testing: 1) to measure the students' mastery of the school's curriculum, or 2) to predict the minimal competence required in the adult world.[3]

The terminology will vary from school to school, and some schools will merge the two concepts by deciding that their curriculum should be based upon minimal adult competence. The different versions of adult

competency tests usually go beyond basic proficiency skills, because they seek to measure an individual's ability to apply 1) basic skills (including literacy) to 2) necessary adult life-role activities such as those of a consumer, producer, or citizen. Examples include ability to understand common indices for comparison shopping, to understand the nutritional ingredients necessary for a balanced diet, to understand a contract for a car loan or home mortgage, to read and understand a newspaper, to fill out a job or loan application, to complete a tax form, to balance a checkbook, to follow a recipe in preparing a meal, to understand proper behavior and attitudes for getting and keeping a job, to use leisure time productively, and to participate as a citizen in the community, state, and nation.

The actual test questions should be subject to close scrutiny, as these examples raise questions about what skills and knowledge are necessary in today's adult world and whether values and background are assumed that discriminate on the basis of race or culture, or that infringe upon individual choice. The law relating to racial discrimination is discussed below. Tests may also be legally vulnerable if they involve mandatory personal and social behaviors that infringe upon an individual's freedom of choice. Furthermore, tests that go beyond basic skills and attempt to measure the affective aspects of social responsibility, good citizenship, self-concept, and job preparedness are vulnerable not only because it is questionable that these "skills" can be measured but also questionable whether they can be generally taught, given the current state of instructional and curricular research.[4]

Where an adult competency purpose is involved, as is the case in many school systems where various adult performance level (APL) tests have been adopted, it is less likely that the school will in fact have taught what is measured by the test. And where this is the case, greater curriculum revision and longer phase-in periods will be necessary. A competency test that measures adult life-role skills that were never taught in the school (and then is used as a basis for denying a diploma) is arguably so arbitrary as to violate due process of law. A competency test lacking curricular or instructional validity might violate substantive due process because then the students are being penalized even though they cannot be personally faulted for poor performance on the test.

Racial Discrimination

While substantial numbers of white middle-class students cannot meet minimal competency standards, there is some evidence that a disproportionate percentage of black and Hispanic students will be adversely affected by the competency test requirement.[5] A number of competency studies show a racial and socioeconomic impact, and this pattern has in fact occurred where competency testing programs have been implemented.

While not opposed to competency testing per se, some black parents in desegregated communities see a racial motive behind competency testing. They say that competency testing was not a major concern at either black or white schools until the schools in their district were desegregated, at which time competency testing was introduced ostensibly "to protect standards." The effect can be resegregation within the

school according to test results (or other forms of tracking), since unequal educational opportunities and/or discriminatory tests may cause black children to score lower than their white counterparts.

Whether or not a racial motive is involved, such practices are arguably unconstitutional in formerly segregated districts. In analogous situations (for example, cases involving ability grouping and voting rights), the federal courts have held that practices that carry forward the effects of prior racial discrimination are prohibited.

The legal standard to be applied to other school districts (those not recently desegregated or found to be subject to prior discrimination) is less clear. As a constitutional matter, the U. S. Supreme Court held in Washington v. Davis[6] that the disproportionate racial impact of a test (in this case, a police department's personnel test) was not sufficient to establish an unconstitutional racial classification without proof that it reflected a racially discriminatory purpose. The Court, however, stated that such disproportionate racial impact can be evidence of discriminatory purpose. And in a concurring opinion, Justice John Paul Stevens emphasized that a person is presumed to have intended the natural consequence of his deeds. Given the numerous studies of the competency levels of public school students (mentioned above), the natural consequence of most competency testing programs will be racial differentiation.

The Supreme Court in Washington v. Davis also found that federal statutory law may provide stronger standards for fair testing than the federal Constitution, noting that the standards for Title VII of the Civil Rights Act of 1964 (prohibiting discrimination in employment) are more stringent than provided by the Constitution, since they

incorporate an effect rather than a purpose standard. Thus when a test or practice disqualifies substantially disproportionate numbers of blacks in hiring and promotion decisions, the burden under Title VII shifts to the employer to validate the test or practice in terms of job performance (that is, to show that the test or practice is sufficiently job-related).

The Department of Health, Education, and Welfare regulations implementing Title VI of the Civil Rights Act of 1964 incorporate a similar effect (rather than purpose) standard, prohibiting practices that have the effect of discriminating against individuals on the ground of race, color, or national origin. Since virtually all public schools are subject to Title VI regulations, competency testing programs having a disproportionate effect on blacks or other protected minorities should be examined in light of Title VI standards, especially where there are indications of unfairness in the test itself or in the administration of the testing program.

Similar standards apply to competency testing programs that have a disproportionate effect on Hispanic children. Some school districts are imposing the competency requirement without a Spanish translation of the test or corresponding curricular and instructional modification. Such practices may not satisfy HEW standards requiring public schools to take affirmative steps to remedy the linguistic exclusion of non-English-speaking children.

The point here is not that minorities should not have to meet the same standards of functional competence as whites. In fact, blacks and Hispanics, who in disproportionate numbers have experienced schools that do not emphasize basic skills, are in the forefront of those

calling for increased attention to basic skills. But care must be taken to assure that competency testing does not lead simply to another form of racial discrimination. The model program provisions suggested below will help to assure fairness in competency testing for all children regardless of race.

Model Program Provisions

The decision to implement a competency testing program raises several far-reaching questions. Should the testing program, for example, be designed to measure only basic proficiency skills such as reading, writing, and computation, or should it go beyond this by measuring a student's ability to apply these skills in adult life-role activities such as those of a consumer, producer, and citizen? Should satisfactory performance on the test, whatever it measures, be a minimum standard to be used in conjunction with other criteria, or should it be the exclusive criterion of satisfactory performance resulting in a high school diploma regardless of age or course credits?

My opinion is that single-criterion evaluation of students, teachers, and public education as a whole sells public education short. For one thing, the state of the art is not sufficiently developed to warrant such exclusive reliance on competency-based evaluations. More important, the primary goals of public education are or should be broader than those reflected by minimal competencies, and students, teachers, and public education generally should not be evaluated exclusively by these narrow measures. Even given a strong commitment to a broader view of public education, there is always the danger that the minimal standards will become maximums rather than minimums.

As in other areas such as drivers' license and bar examinations, however, the need for establishing minimum standards has been demonstrated, and safeguards can be developed to assure that broader goals are not by-passed.

These issues are subject to considerable difference of opinion, as illustrated by early exit programs that allow students, regardless of age, to graduate from high school upon passing a minimal competency test. Competency testing programs obviously involve implicit or explicit decisions about performance objectives and educational goals, and these in turn have important implications not only for curriculum and instruction but also for other school practices such as grouping and discipline.[7] Given the crucial importance of these decisions, a model program would provide for representative community-based participation in the decision-making process.

A model program would also include the following provisions designed to assure that competency testing is fair to students. First and most important, the test should not measure what the school never taught. If, for example, the decision is to test for ability to apply basic skills in life-role situations, satisfactory performance on the test should not be a prerequisite to a high school diploma or grade promotion until the school's curriculum and instruction have been sufficiently matched with these objectives. In other words, the competency test should have curricular and instructional validity, as defined above.

Second, any functional competency test, and the curriculum on which the test is based, should reflect all aspects of our pluralistic society--or at least the extent of diversity reflected by the student population. A functional competency test given in the Miami or

San Antonio schools, for example, should include a number of Hispanic skill and content items, as some cross-cultural competence is arguably necessary for successful functioning as adults in those cities.

Third, a lengthy phase-in period should be incorporated not only to allow for necessary curricular and instructional changes but also to give students adequate notice that failure to learn during this period can have severe consequences. Depending upon the program, this might mean phase-in and notice starting as early as first grade.

Fourth, a model program would provide for multiple learning, evaluation, and remedial opportunities. Students who will have difficulty passing the test should be identified by pretests and given sufficient instruction and/or remediation to help them pass the test. Provision should be made to insure that students who do not pass the test are not tracked in all courses just because remedial instruction is necessary. The program should also provide former students who failed the test the option of further remedial education and test opportunities at any later point in their lives.

Finally, since learning is a two-way street, a model program should provide for students and educators to share responsibility for performance rather than place the full burden on the students. For example, if passing a competency test is one criterion for a student's earning a high school diploma, perhaps a teacher's success in helping students pass the competency tests should be one criterion in the various forms of teacher evaluation (for example, tenure decisions). Such evaluation is feasible if students are tested at the beginning and end of each year in each course and allowance is made for mitigating circumstances such as student absences and transfers. Furthermore,

administrators should be responsible for making the necessary changes in school practices, and the public should be responsible for financing such changes. Developing a workable model of shared responsibility will obviously be more difficult than placing the entire burden on students, but it will also be fairer.

Conclusion

The competency testing movement in this country is gaining considerable momentum, spurred by legislators trying to be responsive to a public call for increased accountability of elementary and secondary schools. Many competency testing programs, however, are unfair to students and undercut public education generally because they do not provide for adequate notice and phase-in periods and are subject to other shortcomings discussed in this article. Educators in particular have a special interest in making sure that any competency testing programs developed in their states and school districts are equitable for all concerned.

Notes

[1]Standards for Educational and Psychological Tests, 1974, p. 28.

[2]See Thomas McDaniel, "The NTE and Teacher Certification," Phi Delta Kappan, November, 1977, pp. 186-88. The South Carolina situation did not involve curricular or instructional validity.

[3]A test with this second purpose should also have predictive validity, or an acceptable substitute. Predictive validity requires a comparison of the predictions about each test-taker based on the test results with the actual functioning of the test-taker at a later time.

[4]See, for example, George Madaus and Peter Airasian, "Issues in Evaluating Student Outcomes in Competency-Based Graduation Programs," Journal of Research and Development in Education, Spring, 1977, pp. 79-91.

[5]For example, Superintendent Joseph M. Carroll reports that when Palm Beach County (Florida) schools first used the Adult Performance Level Test published by American College Testing, 42.5% of minority students failed in from one to five content areas, whereas only 8.5% of non-minority students did so. (See Implementation of a Minimum Competency Skills Program in Palm Beach County Schools, Florida, Joseph M. Carroll and Jerry K. Williams, Palm Beach County Schools, 3323 Belvedere Rd., West Palm Beach, FL 33406).

[6]96 S. Ct. 2040 (1976).

[7]See, for example, William Spady, "Competency-Based Education: A Bandwagon in Search of a Definition," Educational Researcher, January, 1977, pp. 9-14; William Spady and Douglas Mitchell, "Competency-Based Education: Organizational Issues and Definitions," Educational Researcher, February, 1977, pp. 9-15.

A Debate on the Proposition:
Adequate Measurement Technology Exists to
Implement Fair, Equitable, and Useful
Minimum Competency Testing Programs

Carol A. Dwyer
Director, Basic Skills Assessment
Educational Testing Service

As Dr. Nafziger pointed out, the proposition that we have been asked to debate today is, "Does adequate measurement technology exist to implement fair and useful minimal competency testing programs." I intend to demonstrate that it does. The proposition of this debate raises additional questions in my mind such as what is measurement technology, how can we tell how much of it exists, and is the amount that exists adequate? As anyone who's ever tried to answer that adequacy question knows, it carries with it its own baggage of prior questions and I'll address some of those a little later on in talking about standards. In my role as director of a developing minimal competency testing program, the Basic Skills Assessment, the status of measurement technology in a number of areas had to be assessed and questions were raised both by the staff members at ETS and by the local school district personnel who sat on the policy board of Basic Skills Assessment. We looked at the adequacy of technology in areas such as defining the skills to be tested, in program planning, in objective test planning and development, in establishing scales and interpretive data, in results delivery systems, and especially, in standards setting procedures for local school districts to use. In none of these areas was the adequacy of technology question more prominent or more troublesome than

in the area of setting standards. I'd like to concentrate on that area of setting standards in my remarks this morning.

Performance standards form the crux of a fair and useful minimal competency testing program. They represent the ultimate point of data reduction. When a school district or state has instituted a minimal competency testing program, the bottom line for schools or for students is "has the standard been met?" Basic Skills Assessment initiated this series of inquiries into standards setting procedures which resulted in our taking a very positive stance toward the setting of standards and also, in our assertion that the process can be undertaken with ease by non-technical school district personnel. I want to emphasize that I'm not trying to minimize the complexity of the task of setting standards; only the complexity of the technology required to carry it out. I believe that the measurement technology as it currently exists is more than adequate to set standards. The practice requires only very elementary statistical concepts and manipulations. Criticism of the technology of setting standards often focuses on the role of judgment in that process--on whether judgment is a necessary part of standard setting and on how much judgment represents an acceptable level in the standard setting process. Judgment has taken some strong abuse from people in education as well as from the hard science types and that's not a recent phenomenon. Critics of the role of judgment in standard setting often equate judgment, subjectivity, values-oriented decision-making with arbitrary and capricious decision-making. I'm trying to stress today that a standard setting procedure which includes a number of explicitly stated judgmental steps can produce fair and useful minimal competency standards.

All methods of setting standards involve judgment. Certainly all of the ones that Basic Skills Assessment urges people to use require judgment. This is not just a deplorable state of affairs due to our current state of ignorance. It's a necessary and desirable intersection of social values and measurement technology. The setting of standards is as much a social and political process as a psychometric one. Even if highly sophisticated statistical procedures were performed flawlessly, the standard setting procedure would still be deficient if the social and political aspects of the process were excluded. In other words, judgment doesn't contaminate the standard setting process, it strengthens it.

I maintain there are two principal aspects to that strengthening. The first is that in a sound judgmental process, assumptions are made explicit. In comparable assumptions in statistical procedures, the assumptions are either not made explicit or are met only to a limited degree. Second, judgmental procedures address whole areas that statistical procedures cannot, such as the social, psychological, and policy implications of setting a standard at a particular level. We've recommended that those who set standards in the Basic Skills Assessment view the setting of standards as a long and complex process in which the phase of data collection and analysis occupies only one niche. We've stressed the importance of the social and political context of standard setting and say that awareness of the potential effects of the standard setting process should influence the judgments that enter into that process.

Here are some of the requirements for an adequate standard setting process: Involve all of the stake holders or constituents in the process.

Those making judgements leading to the standards should include representatives of both sexes, of racial and cultural minorities, and of the interested parties such as school boards, students, parent-teachers associations, and the staff of the schools that will be affected by these standards. Two, make a public commitment to the process and to the standard. The more open the process, the less likely it is that a standard will be arbitrarily rejected after it's set. Three, use standard setting as an iterative process, producing a series of tentative estimates. This implies a formal review of the standards at least annually and also assumes that the experience gained from a standard setting process will be useful in future iterations of the standard setting procedures. Four, allow sufficient time for the process and allocate sufficient resources for it. Plan a small scale pilot study for the standard setting process and review the results before you do the real thing. Basic Skills Assessment specifically recommends using one or more of four methods of setting standards--two of which involve judgments about test content and two of which involve judgments about the mastery status of the students who are taking the tests (Zieky and Livingston, 1977). I'll just describe these four methods briefly so that you'll have some familiarity with them for those of you who don't know them by name.

The first of the two methods that are based on judgment about the test content is the Nedelsky method, which is based on collected judgments of which distractors in multiple-choice test questions a minimally competent student could be expected to eliminate. Pooling the resultant probabilities of a correct answer yields a sum of items expected to be answered correctly by the minimally competent student.

A second method based on inspection of test materials is the Angoff method. The logic of this method is the same as that of the Nedelsky method but the questions are considered in their entirety and judges are asked to make assessments of what proportion of a group of minimally competent students would be expected to answer this question correctly. I should point out that the Angoff method has the advantage of not necessarily being restricted to multiple-choice test questions. The third method is one that involves assessing the mastery status of students. We call that the borderline group method, and it involves simply asking for the judgments of knowledgeable persons, usually teachers, about the mastery status of individual students. Those students who cannot clearly be designated as either adequate or competent (or, on the other hand, inadequate or not minimally competent), form a "borderline" group. The median of that borderline group is then taken as the point at which the standard is set. The fourth method is the contrasting groups method. This method is also based on collecting judgments about the mastery status of students. In the contrasting groups method, students are classified either as masters or as non-masters of a particular skill. Separate frequency distributions are generated for each of those groups, and the point of intersection which defines either the least number of misclassifications or the appropriate proportion of misclassifications in either direction is taken as the point of the standard. Additional value judgments required by these methods include first, an agreed-upon definition of minimal competency--a definition of what is basic and how you tell if you have a minimally competent student or not. In the contrasting groups method, an additional

judgment is made concerning the likely consequences of different types of misclassifications or classification errors.

To sum up, I believe that the technology that I have described and the technology underlying other methods of setting standards is more than adequate to implement fair and useful testing programs. But it is not adequate without some help from the group process required to set the standards at a level acceptable to all the constituents; from a spirit of compromise; and from common sense.

Reference

Zieky, M. J. & Livingston, S. A. Manual for Setting Standards on Basic Skills Assessment Tests. Princeton, New Jersey: Educational Testing Service, 1977.

Loretta A. Shepard
University of Colorado

My position is that all standard setting procedures that currently exist and, in fact, that are likely to be developed rest on human judgment and hence are arbitrary. Perhaps they are carefully arbitrary but they are nonetheless arbitrary and the main point of my argument is that because of that arbitrariness, there will be large differences depending on where the standard is set in the percentage of students who fail and the percentage of students who will be passed. To buttress that argument, I would like to refer you to some good resources, some excellent reviews of known standard setting methods. Those sources include a fairly old article by Jay Millman (1973), and a more recent and more extensive article by Meskauskas (1976). I think that the Basic Skills Assessment manual for standard setting (Zieky and Livingston, 1977) provides a very down-to-earth discussion of the standard setting methods, and in addition there's an article by Jaeger (1976) that goes into some of the internal validity problems of various standard setting methods. All of these authors conclude, as did Carol, that standard setting procedures require judgment. I'd like to elaborate a little on some things she's already said.

One of the best methods among those that require examination of test content to make a decision is the Nedelsky method, which she mentioned. It is eminently sensible because it provides a systematic way for the judges to examine the test content item-by-item and come to a conclusion about the minimally competent student. Basically what the Nedelsky method does is rule out the distractors that the minimally

competent student could rule out and then derive a chance score for the remaining distractors. So, if you can rule out one answer and have to choose from the remaining three, your chance is one in three of getting the correct one. The judges just accumulate those chance scores from all of the items until they have the minimal passing score. It's better than some of the other methods because it takes into account the distractors as well as the correct answer, which is what the Ebel method takes into account. The Angoff method is suggested for those instances where you don't have a multiple choice format and so you just sit down and reason, "Hmmm, I think this looks like a .4 or a 60 percent item." So you reason your way to the probabilities rather than going through the distractor method.

I'd like to illustrate the method of identifying groups that Carol was talking about. If you're not judging test content (and those that I've just mentioned rely on judgments about test content), then another thing to do is to look at groups of known masters and non-masters and decide on the optimal cutting score that will help you distinguish between those masters and non-masters. As shown in Figure 1 it's basically a normal distribution of non-masters and a normal distribution of masters and we have a cutting point at the point of overlap that minimizes the two kinds of errors--the false positives and false negatives. The problem with this method is that judgment is relied upon to identify the masters and non-masters. Sometimes that is made easier by just casting out of the group of study all those that we're not quite sure about. So we'll just find the masters and non-masters and if we're not sure about an individual student, we'll leave them out of the group. But that kind of wishywashiness is exactly what moves the center; it

changes the overlap and it moves the center back and forth so that the cutoff scored could be different enough to allow very different percentages of students to pass. The ETS group has come up with an alternative suggestion which focuses on the group we're not decided about--the borderline group as they call it. (That's these people someplace in here--the gray area between masters and non-masters.) We just identify all of those that in the previous method we were unsure about and use their median on the test as the cutting score. Now, that's a handy decision but why not the median plus 10 percent or the median minus 10 percent? Each of these will make a difference in the number of students that pass.

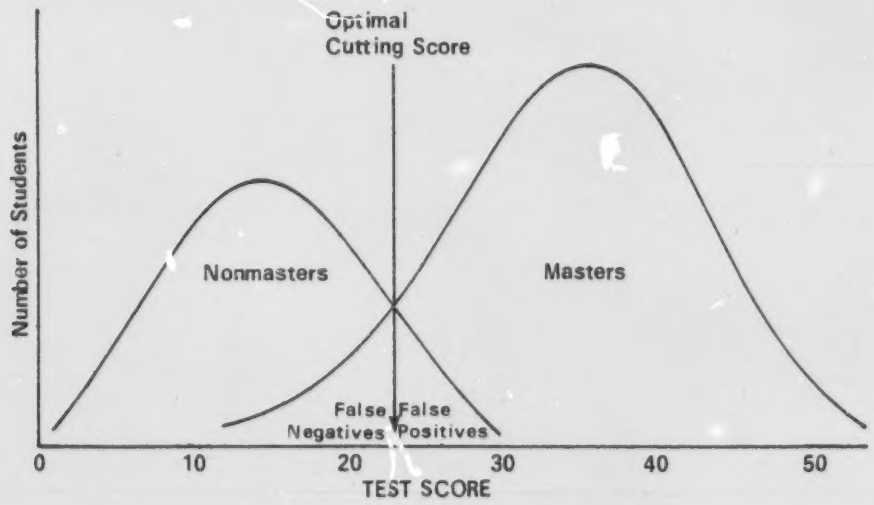

Figure 1. Illustrative of standard setting by selecting the cutting-score that minimizes the overall number of errors in classifying masters and nonmasters.

The most elaborate methods of standard setting, the decision theoretic approaches, elaborate on this basic method by saying, "We'll also take into account the cost of the two kinds of errors." The decision theoretic approach is an improvement on some old work by

Livingston which suggested that a least squares solution to the problem will minimize the two kinds of errors. And their conclusion now--Hambleton and Novick's (1973) more recent conclusion--is that using the least squares technique over-emphasizes large deviations on the wrong side of the cut score. In other words, it weights more heavily scores here or scores there and we're just concerned with whether they're on the wrong side of the cutting point, not how far to one side of the cutting point they are. So that's what they have termed the threshold loss problem, the threshold error problem. They suggest that we minimize the two kinds of error of classification. If we can also assign different utilities to these two kinds of error, we will differentially weight them (and come up with a function or formula that describes those kinds of error and their relative weights and optimizes the decision) so we can minimize the kind of error that is considered more costly. All of this requires that we sit down and reason about how serious it is to reach a false positive or a false negative conclusion. The whole procedure, as do the content inspection procedures, rests on how wise we are. The only thing that is wrong with human judgment (and we have heard advocacy that human judgment is the strength of these procedures) is that humans make mistakes. And in this case, humans may make mistakes that will be costly to other humans.

I'd like especially to call your attention to a more recent paper by Ron Berk (1976) because he says his method doesn't rely on judgment. What he's proposing is basically like the illustration in Figure 1 where we identify masters and non-masters. But he thinks he's avoided the judgmental problems associated with that method by defining masters and non-masters as instructed and uninstructed groups. So, naturally the

instructed group will have mastered the task and the uninstructed group will not, and we have an empirical way of making these two distributions and don't have to rely on fallible human decisions. The problem with that is how long should they be instructed? How long do they have to be instructed until we'll consider them masters? And where are you going to find the uninstructed group? Well, we could get some two-year olds in there but the problem is we want an uninstructed group that is like the group who will take the test in all respects except that they are not masters.

I bring up this example just because I'd like to leave no doubt that all of these methods rely on human judgment and that we can make mistakes when we rely on our wisdom to arrive at these cutting points. My advice is that we should avoid making such judgments whenever we can. I will also come to the conclusion (and I've written a paper that sounds like Carol's suggestion) that when standards must be set, it should be done by an iterative procedure, involving all of the audiences, and allowing for redress when the standards seem to be inaccurate.

I agree that we must come to that kind of solution when we are forced to set standards. But why can't we suggest that in many instances the problem or the motivation behind the request for minimum competency testing can be addressed in other ways? For example, if the reason for minimum competency testing in your state is the SAT score decline, for heaven's sake, let's remind the legislators what you in this audience already know, which is that the people who are doing less well on SAT tests are not the ones who stand at risk on the minimum competency test. It's an entirely different population and I would propose that you might better address the SAT score decline by

reinstituting requirements in writing and literacy or more advanced literacy--even literature--in the high schools. Instead of having rap sessions we could teach writing or we could teach advanced literature. And that's a more direct attack on the problem.

If the problem is the more compelling one--that there are instances of students who graduate from high school who cannot read and may in fact even sue the school district because they cannot read--then I think there are other procedures for making sure that somewhere earlier than as a senior in high school, we discover that students are not making the grade and need remedial instruction. I propose the same kind of thing that was mentioned this morning with regard to special education in individualized education programs: we consider that it is now necessary to make more formal the requirements by which teachers, and many times parents and administrators, decide what will be done about an individual student. And I submit that the only important thing about such a program is to make sure that the child doesn't get all the way to being a senior in high school without having learned to read <u>if the school could do something about it</u>. That's the important phrase. <u>If the school could do something about it</u>, we want to have identified that child and what the school can do for that child at some earlier time to have made a difference. It might be staffings like those that are used when we make decisions about special education youngsters, many of whom are the same youngsters that we're talking about perhaps failing these tests. Those suggestions would be eminently more sensible than very costly--in dollars and in human misery--testing programs.

There will be instances when we cannot get out of a minimum competency testing program and we have to do the most sensible thing, that is, set standards that are based on reason where people have been brought together and provided with as much information as possible.

One further point that I have to make, however, is that I cannot think of an example in which we will feel compelled to set standards for making group decisions. I can understand the motivation behind legislators and educators who are saying we will examine individual students and decide whether they should have a certain kind of diploma. But I cannot understand why educational programs need to be evaluated in terms of performance relative to a standard, because it is invariably true that differences in performance on sub-tests or tests will be due to differences in the stringency of the standards rather than true differences in performance. I recommend to you a paper presented by Glass at the mid-winter conference at UCLA about the situation in which Florida students are doing so much more poorly in math than in reading. This is what I mean, this is an example of a group decision, a conclusion about the school system that uses standards rather than some kind of basic normative comparison. I do not advocate comparison to norms directly but I do advocate inspecting relative differences in comparison to norms as a preferable way of identifying strengths and weaknesses in programs, rather than so-called absolute standards which we know are arbitrary and which we know are fallible. By the normative comparison, I mean that if students are relatively better off in relation to the norm in reading than they are in math then I would be convinced that math is a problem in that school system. But I am not convinced that Florida ought to spend all of its remedial money, that extra $10 million,

on math rather than on reading and that's what they propose to do. That money is earmarked for math because that was the subject that more students failed.

If we must then, for individual decisions only, have standards, then let's proceed without any sense that there is magic or accuracy in the scientific mumbo-jumbo that has been offered to us. I'm not aiming that at ETS because, as I said earlier, the Zieky and Livingston manual is very down-to-earth and very sensible. But we have a lot of pretty fancy standard setting models that often lead people to believe that there is a scientific precision there that does not really exist. If we must set standards, then I think that we should give as much information as possible to the standard setters, like normative information and like projected estimates of how many students will fail if they set different standards. I think it's as important to consider how many will fail and how many will pass by different standards as it is to inspect the content of the test and say, "Hmmm, that looks like a pretty good number of items."

References

Berk, R. A. Determination of optimal cutting scores in criterion-referenced measurement. Journal of Experimental Education, 1976, 45, 4-9.

Hambleton, R. K., & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Jaeger, R. M. Measurement consequences of selected standard-setting models. Florida Journal of Educational Research, 1976, 18, 22-27.

Meskanskas, J. A. Evaluation models for criterion-referenced testing: Views regarding mastery and standard-setting.. Review of Educational Research, 1976, 46, 133-158.

Millman, J. Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 1973, 43, 205-216.

Zieky, M. J. & Livingston, S. A. Manual for Setting Standards on Basic Skills Assessment Tests. Princeton, New Jersey: Educational Testing Service, 1977.

The following exchange between Carol Dwyer and Loretta Shepard concludes the debate.

CD: Well, I'd like to take issue with a number of things you've just said. First of all, at the very nitty-gritty level, concerning your comment about not being able to imagine the situation in which we'll be compelled to set standards for groups of individuals, I think we're seeing that right now. There are people all over the country being compelled by non-measurement, non-education people to conform to, even in the worst situation, a standard that has already been set for them.

LS: I'm sorry, I didn't say that I couldn't imagine when we'd be compelled to do it for individuals. I said I can't imagine when we would have to do it as the only way of evaluating a school program. So, I think we can get out of the group problem and I think we ought to get out of and in fact dig in our heels if ever forced on the group problem. I agree that we cannot get out of all of the problems associated with setting standards at the individual level and that's where these last ditch and sensible requirements, given the imposed requirement, are recommended. But for groups, I think we can get out of the need to set standards.

CD: There are a couple of things you mentioned in the course of demonstrating that judgment is arbitrary that I'd like to comment on. We had some interesting work in the course of Zieky and Livingston's developing procedures for the Basic Skills Assessment. One of the things we noticed was that the judgment of masters and non-masters could be structured by directions given to the people who were doing the judging. We

had a series of tries at defining mastery, providing people with different definitions of masters and non-masters and found out that some work conspicuously better than others in terms of getting a coherent picture of mastery/non-mastery. Of course it is judgment that moves the cut point around, as you demonstrated. The thing that leads to moving that point is ultimately how stringently or how leniently people classify students as masters or non-masters--how high their standards are in a particular situation. As you may expect, I consider that a good thing rather than a bad thing because that puts the standard at a sensible place for that school district. We are talking about people who need to make, in most cases, a practical decision either for themselves or to satisfy somebody else. When you have set your cut point based on lenient or stringent judgment, that puts it very close to your real life values. The school systems that have able students and are competitive have high standards; they have stringent standards that adequately reflect their political and social values.

LS: All right, I have something to say about that. First of all, there is good evidence--not a lot of evidence--but there is good evidence that judges don't agree. In the Meskauskas review, there is evidence that judges using the Nedelsky method who were trained to use the Nedelsky method, set cut-off scores that varied as much as from 38 percent to 80 percent on multiple choice items. Now that's in percentage of items required to pass. There is also the Andrews and Hecht (1976) paper which demonstrates very serious differences when we compare the Ebel and Nedelsky method. If your retort to that is "Well, that's okay because they really have different conceptualizations underlying those two rules and naturally they'll be different," they also present compelling data

that two sets of judges using the Nedelsky method came up with very different results.

CD: I'm sure those people had very different standards for what represented minimal competency.

LS: And that's my other point. The problem is people in Palo Alto don't stay in Palo Alto. Don't you know, haven't you read the Time magazine articles about people fleeing from Florida to Georgia because in Georgia they presently don't have to pass a test to get a high school diploma? The same thing is true between counties and districts. If you happen to live in Beverly Hills, which is a tough district, you can just flee to central Los Angeles and pass a different test. We have a mobile society and we're going to have these people rushing about because the standards are so seriously different that in one district you can pass and in another district you can't. Now, I'm not advocating by any means national standards or state standards but, for heaven's sake, let's wake up to what will happen if we allow real differences in what the standards are. Now, I just can't believe there should be these kinds of differences in what it takes to be a minimally competent citizen--we're not talking about being ready to go to college. If you wanted to have a standard in Palo Alto because you expected those students to go to college, then I can understand that what you taught in that high school would be different. But we're talking about minimums here. And if there's any sense about minimums, I think they should be the same.

CD: Well, I don't think so. You're taking a definition of minimal competency that would be meaningless to Palo Alto.

LS: I don't think Palo Alto should give such a test.

CD: How about if Palo Alto gave a test that was at the appropriate level for the minimally competent to graduate from a fancy high school like Palo Alto's and go on to college?

LS: I think that this is very troublesome ground. We have the SAT test to give to students who want to leave Pali High and go to college. We have entrance requirements to college and we also have entrance requirements for certain jobs that now, by law, must show the relevance of those requirements for that job. I don't see that the public schools have to take on that certification requirement--for one thing because they're supposed to be special for every job, because you're not supposed to give a test that is indirectly an ability test and withhold a certain job because someone cannot pass the test. It is by law that you must now demonstrate the relevance of that test to that job.

CD: Well, to get back to the original point, if everyone is in agreement or made to be in agreement with a single definition of minimum, in what way does that differ from a national standard?

LS: Well, you can't catch me in that position because I'm just showing you how futile it is to try to set standards. You can't say that I now must support the position that we ought to have a national standard.

I do have one other point to make in response to Decker Walker this morning. I think that we all ought to cringe at the fact that where we had assessment programs previously, that might have been able to detect the effect on curriculum of the imposition of minimal competency testing, we now have states who are abandoning their old

assessment program to take up the flag of minimum competency testing. In New Jersey they are no longer going to test the things that they now concluded by committee are not minimums. So we will lose in those states the opportunity to monitor what happens as a result of minimum competency testing. I feel like I'm fighting a last ditch battle here and I'm going to have to live with minimum competency testing over my dead body. And if you want to do it over my dead body then what I insist on are the kinds of procedures that Carol was talking about to make them as explicit rather than implicit as possible. And then, for heaven's sake, please leave me the vestige of an assessment program or some kind of evaluation program so I can demonstrate five years from now what happens to curriculum and what happens to student performance as a result of these new testing programs.

CD: That seems sensible. I'd like to raise one other question. I have just one more and maybe it's a short one, because I think there's a quick answer to it. You are asking about or recommending the procedure of looking at improvement as the only way of determinimg progress--looking at yourself this year in relation to national data, and looking at yourself next year and seeing if you got better or worse. How much better or worse do you have to be? Would you set a standard on that improvement and if so, how?

LS: What I was referring to is not so much between years, though that makes sense as well. I was talking about if you want to see whether you're relatively better off in math than in reading, I suggest that you look at your position relative to a national norm. Better still, if you have prediction information based on background variables in your

district, look at your relative position compared to an expected score for that normed test. That will show you whether there are serious differences. In that case, I trust the statistical standard which is to allow for measurement error and then see how serious the discrepancies are. But I can still say big differences are worse than small differences. Unless they are very large, certainly larger than could be explained by chance, I'm not going to allocate more money to math than to reading.

CD: Well, that seems sensible too. I just want to dredge up a quote from Michael Scriven to finish off and have my last word. He said in a presentation at an ETS Invitational Conference that the way to make hard decisions with soft data is in general not to harden the data, although that could help, but it's to improve the decision-making process. I'm hopeful that that's where common sense will prevail in minimal competency testing in the near future.

PART TWO

PANEL REACTIONS

Escaping the Minimum Competency Dilemma:
Scaled Certification

Ellis B. Page
University of Connecticut

As everyone may know by now, American education is in something of a stew about minimal competency. On the one hand, we have "meaningless, disgraceful" diplomas issued by local systems to twelfth graders, some of whom, alas, cannot read past a fifth grade level. What is the use, people ask, of such diplomas? Citizens, taxpayers, parents find such lack of standards an outrage, and state after state is trying to establish some floor of guaranteed intellectual ability before issuing a document of high school "graduation." Many lawmakers have happily seized on this as an obvious, popular issue. And many educators, less happily, believe that minimal competency testing (MCT) is inevitable (Shepard, 1976).

On the other hand, once we have assented to MCT in principle, we find ourselves mired in a nearly incredible swamp of problems: technical, administrative, legal, political (Miller, Graham, & Hill, 1978). Consider the political: When "respectable" standards are applied to real populations, they screen out perhaps one quarter of the potential graduates. Yet no one believes that such a proportion would be acceptable in any particular school, let alone for a district as a whole (Brickell, 1978, p. 9). Consider the legal: There is a grave question whether it is permissible, after twelve years of "passing" a student through school, to deny that student certification at the end (McClung, 1977, 1978). And what of the "discrimination" which will be charged when the failing students show, once again, a high proportion of certain minorities?

Consider the administrative: What exactly shall be done with those who have "failed" to meet the prescribed standards (Walker, 1977)? And who shall prescribe the standards? And consider the technical: As measurement specialists, many of us find MCT a carnival of horrors, centering around that strange, fictional line which is used to separate the "minimally competent" from the "incompetent."

Thus, we seem caught in a dilemma without exit. It is intolerable, from some perspectives, to go on issuing trivial diplomas which cannot be trusted by other schools, by employers, or by society. But it seems equally nonsensical to blunder into a new era of great turmoil, ceaseless litigation, and (probably) still more court-made educational decisions. Is there a way out of this dilemma, avoiding both of the unsavory alternatives which are currently debated? After some study, I believe there is.

The Irrational Dichotomies

Despite controversies in the popular and professional media, specialists in measurement can probably lay claim to the best body of theory in the social and behavioral sciences. Yet the establishment of good test theory and practice is a recent accomplishment, marking the past several decades, and coming long after the hardening of our schools and colleges into their present, general shape. The educational system, indeed, seems based upon the kind of yes-no, categorical thinking which dominates the current arguments about MCT. If a student seemed minimally competent at one level, then the promotion, or diploma, was awarded. If not, then it was denied. When it was recognized that there were still enormous differences among the "competent," the schools and colleges established "honors." But these, too, were categories, new dichotomies

troubled by the same fuzzy borders and arbitrary decisions.

Fact is, in measurement we have long outgrown this categorical view of human abilities. All of our evidence (and it is enormous) argues quite a different picture: Mental abilities, resulting from both genetic and environmental influences, are continuous and normally distributed. They range from the astonishingly gifted to the depressingly dull. Specific educational accomplishments ("competency" areas, if you like) are best seen as traits falling in much the same pattern (Page, 1976), with most students clustering close to "average" performance, but with extraordinary individuals trailing off in either direction. Typically, these specific traits are correlated with each other, a fact strengthening our belief in g, a general intellectual ability.

Now let us consider one such specific trait of fundamental importance: reading. All standardized reading tests, though developed independently, correlate highly with each other. These will do an excellent job of rank-ordering individuals (whose scores are markedly different) according to how well they can read a wide range of materials, whether from school assignments or from "real life." In that adult life, many jobs will require and reward such reading abilities. But some, speaking comparatively, will not. Today, many professionals seem determined to debate how, in such a world, we can draw lines on which scores will mark those "qualified" for countless unspecified roles as employees, citizens, or parents.

The answer seems to me unmistakable: We can't. It is not simply that we can't draw such a line technically. We can't draw it in principle. And it is a great time-wasting, resource-consuming self-deception to imagine that we can. In this light, though sympathizing with the

frustrations which have led to the MCT programs, I have little optimism about their future, as they are presently outlined. Indeed, I would even drop the language of "minimal competency," since it describes a concept which cannot be operationalized or applied.

Scaled Certification

But such intractable dilemmas result, very often, from outgrown or poorly expressed theory. Suppose we recognize the validity of the charges against meaningless diplomas. But suppose we sidestep the difficulties with cutpoints and, instead, use a more accurate and theoretically defensible way of expressing a student's accomplishment. We can do all this, without any very radical changes in process, by simply writing on the diploma the results of a test battery in the most important subjects. We can call this sort of diploma scaled certification (SC). I believe that SC can be structured in such a way that it overcomes the most telling objections that can be raised against it.

Once proposed, it may seem the dullest orthodoxy. And it has, indeed, a well-worn, established parallel in our college-admissions testing of the present day. Every year, millions of high school juniors and seniors take such examinations, most often the Scholastic Aptitude Tests of the College Entrance Examination Board of New York City, or the American College Testing Program of Iowa City. Both SAT and ACT have had a steady, incremental growth in use, and now clearly are important national institutions, depended on by colleges, high schools, and students to help with this extraordinary complex sorting out which goes on every fall.

Among the most remarkable features of the SAT and ACT is their relative freedom from scandal. With the huge rewards dependent on such tests--educational, vocational, financial, it is remarkable that no major,

extensive breach of security has damaged either program. And their procedures and safeguards should be carefully studied in any establishment of scaled certification.

Clearly, neither SAT nor ACT provides a perfect model for any national SC programs. For one thing, they have been designed to discriminate among the upper half of high school graduates, by and large, rather than across the range of talent. For another, they are conceived of as describing abilities, rather than achievements. Yet technically, neither of these reservations is a very practical one. And in general, an SC program might be quite like the ACT, with its four areas of English, math, science, and social studies.

How detailed should such SC be? Once we abandon categorical thinking then four or five measures, similar to those of the ACT, may be about as far as we should go. We surely do not want to fall into the trap of splintering behaviors into many unreliable sub-tests, such as spelling, knowledge of local government, or calculating percentage. In fact, we probably should avoid the goal of specificity which marks the "criterion-referenced" movement, and stick pretty faithfully to the norm-referenced, or domain-referenced, concepts which have proved serviceable in standardized test programs in general.

The apparent advantage of norm-referencing, for such SC diplomas, is not philosophical, but practical. Just as we cannot decide a cutpoint for "minimal competency," so we cannot decide for any important trait, a point at which the student "can read," or "can do algebra," or "knows U. S. history." For such accomplishments, we can best express abilities in the traditional, comparative way. And decision makers will learn soon enough, just as colleges have for admissions-test scores, what the SC

achievements will promise in terms of performance in life, or in further schooling. For some there will be high relevance, for some not.

Some Questions and Answers

A number of questions come up in the consideration of any major program, and deserve attention here:

1. In just what way is scaled certification better measurement than minimum competency? Once we acknowledge that these traits are continuous, then the observed scores reported by SC diplomas will be much closer to the underlying "true" scores than will any gross categories.[1]

2. In what form might SC scores be reported? Personally, I would lean toward the T-scores, with mean of 50 and standard deviation of 10. Of course, such diplomas should be accompanied by interpretive material, of the kind which now accompanies SAT or ACT.

3. Will students be "branded for life" by such scores? Not at all, anymore than they are now by standardized tests. So far as retests are concerned, I would favor, once again, the college-admissions models: tests given possibly four or five times a year, should students wish to improve their record.

4. How would we handle different test results by the same student? What some testing companies do is simply to report all scores made at any time. But this seems to me too cumbersome, and inappropriate for some purposes. I'd favor a report that is weighted toward the more recent testing. For two trials, the second might count twice the first. For three trials, the third might count as much as the first two combined. And so on.[2]

5. A profile of scores is often too hard to handle. Can we come up with a single score? Yes, of course. The ACT, for example, has a

composite, as do most other achievement test programs. To give each score its proper value, I would suggest a weighting plan like that we explored for the "bentee" (Page, 1972, 1974; Page & Breen, 1974).

6. How does SC handle the problem of low-scoring minorities? It avoids the problem, by avoiding any decision-making. This seems a proper stance for school systems in general to take right now, when the vexed questions of "affirmative action" vs. "reverse discrimination" are being thrashed out in the courts (cf. Bakke and Larry P.) and in public opinion. It would be consistent with the reasonable, current practice of reporting scores as they are earned, and leaving it to others to use these in decisions, for minority and majority students alike.

7. And how does SC overcome the current technical, administrative, and philosophical problems of minimal competency? It does not pretend that all high school graduates are equally able, or equally educated. In this sense, diplomas using SC would no longer be "meaningless" sheepskins, clothing the brilliant and the illiterate with the same empty classification. On the other hand, scaled certification would not trap educators in the impossible task of drawing arbitrary and unstable lines, and then defending them against a host of attacks.

In summary, we don't need to resign ourselves to the present, hollow diploma, nor to the endless difficulties of a minimal-competency future. Scaled certification--the reporting of a brief profile of scores from well-designed senior tests--may be an honorable way out of the dilemma, a fairly conservative program we can live comfortably with in the decades ahead.

References

Brickell, H. M. Let's talk about minimum competency testing. Denver, CO: Education Commission of the States, 1978.

McClung, M. S. Are competency testing programs fair? Legal? Phi Delta Kappan, February 1978, 397-400.

McClung, M. S. Comptetency testing: Potential for discrimination. Clearinghouse Review, September 1977, 439-448.

Miller, B. S., Graham, P. A., & Hill, W. G. Minimum competency testing: A report of four regional conferences. St. Louis, MO: CEMREL, Inc., January 1978.

Page, E. B. Seeking a measure of general educational advancement: The bentee. Journal of Educational Measurement, Spring 1972, 9 (1), 33-43.

Page, E. B. "Top-down" trees of educational values. Educational and Psychological Measurement, 1974, 34 (3), 573-584.

Page, E. B. An historical step beyond Terman. In D. P. Keating (Ed.), Intellectual talent: Research and development. Baltimore, MD: Johns Hopkins University Press, 1976, 295-307.

Page, E. B. & Breen, T. F. III. Educational values for measurement technology: Some theory and data. In W. E. Coffman (Ed.), Frontiers in educational measurement and information processing. Boston: Houghton-Mifflin, 1974. Ch. 3, 13-30.

Shepard, L. A. Setting standards and living with them. Florida Journal of Educational Research, 1976, 18, 28-32.

Walker, D. F. Editorial: The hard lot of the professional in a reform movement. Educational Leadership, November 1977, 83-85.

Notes

[1]This is not the place for extensive test theory, but it is obvious that the squared deviations of the observed scores from continuous, hypothesized true scores will be less when observed scores are themselves reported in more continuous form.

[2]The suggested plan would score the student, where multiple attempts have been made, by the formula:

$$C = \text{Integer}\left[\left(\sum_{k=1}^{n} kX_k \right) / \sum_{k=1}^{n} k \right],$$

where C is the cumulative score in a trait, measured across time; k is the counter of the number of times (n) the student has taken the test; and X_k is the score made by the student on the kth trial. Where there is a considerable interval without retesting (say three years), it might be possible to establish a new score on one trial.

Comments on Minimum Competency Testing

Ruth S. Nickse
Nickse Associates Incorporated

I think you're probably overwhelmed by the enormity of the issues presented here today. The panel has done an excellent job of giving minimal competency testing a smack between the eyes. There is a need to summarize some of the problems that have been expressed here today, and a need to reiterate concerns expressed by both proponents and opponents of minimum competency testing.

Perceived Negative Effects

Pressure

The origin of the pressure for MCT is one problem that concerns us all. The mandate by states--the coerciveness that seems involved in MCT--makes people from the education world uneasy. Although there are good reasons why constituencies and legislators are lobbying for it.

Values and Standards

The values question is very severe. As Decker Walker asked, who chooses, what kinds of competencies? How do you get consensus? And, as we can tell by the debate here, the standards and levels of performance question is a very ticklish one.

Costs

There is a question about costs that hasn't been mentioned much today, but has been written about in a paper by Barry Anderson (1977). Perhaps the financial price is too high to develop adequate and reliable tests. Will we sweep toward national standards and tests because it is

less expensive than a state-by-state effort? Can states afford good MCT?

Testing Problems

As discussed here today, there are innumerable psychometric problems. The difficulty of designing valid tests that confirm skills and respect cultural ethnicity is one major problem. There is also the danger of labeling kids as incompetent at 13, or stigmatizing kids who can't seem to get that diploma for graduation.

Credentialing

There is a credentialing problem as well. Because competence levels are attached to graduation requirements, there will be a danger of mass failures. Inasmuch as we are a credentialled society, this condition leads to lessened access to further education and jobs.

Effects on Instructional Content

A possible trivialization of the content of instruction might occur because we may minimize our expectations and thus minimize instructional objectives to teach to the test.

Remediation Costs and Psychological Effects

We are not paying attention to where the money is coming from for remediation. How are we going to target our dollars for instruction in areas in which we find kids are weak? And we have the problem of the special kids--the bilingual and the handicapped kids. What's going to be the effect of MCT on them? Some schools, as you know, are beginning to hold back on social promotions and are asking that the children meet minimum requirements for passing from grade to grade. What's the psychological effect of grade retention on them?

Effects on Instructional Staff

The taxing of the instructional staff, particularly at the secondary school level, is going to be great. I get a kick out of this argument that it is the high schools that are at fault for the low performance problem we seem to have across the country, when basic skills, as we all know, are generally taught in the elementary schools. Somehow or other, the elementary schools seem to be escaping the heat. They should be sharing the responsibility. These are some of the concerns which mitigate "against" competency testing.

Some Benefits of Minimal Competency Testing

There are some possible benefits from the minimal competency testing movement.

Emphasis on Community Input/Responsiveness to the Community

Generally, minimal competencies are identified by some process which involves constituency groups from the community--including taxpayers, parents and employers who hire persons at the high school level. Such participation encourages responsiveness and creates an atmosphere of cooperation. It reflects a concern for all manner of public interests: accountability, basic skills, and education for the handicapped.

More Meaningful Credentialing

The credentialing process and graduation standards will be sharpened up by this practice. It should make clear what a district's high school diploma means. The worth of this depends, of course, on what standards, levels and competences are chosen.

Clarification of Content/Strategies for Teaching

The identification and selection exercise which establishes a district's list of required or desired competencies presents an opportunity to examine the nature of general education. This clarifies instructional content. Besides clarifying what is taught, strategies for teaching and evaluation are also clarified.

Costs

Targeting of scarce resources (staff time, tax dollars, etc.) will be, theoretically, easier to accomplish.

Affirmation of Skills

Attitudes toward learning might be changed with minimal competency testing. There is something to be said for an affirmation of skills, and competency testing can be used in this fashion. For students, a good score on such a test sometimes leads to self confidence and motivation to continue learning. I'm thinking of an example from the California Proficiency Examination. As you may know, California has an early exit exam called the California Proficiency Exam, which is administered several times a year. There was a fear that students who passed this exam at age 16 would be leaving the high schools in droves. That fear is largely unfounded and numbers of kids are taking the exam but not dropping out. The test acts as a standard by which students measure themselves and check their own progress.

Improved Classroom Management

Classroom management could be improved by minimal competency testing, because it could make performance agreement possible both within individual classrooms and throughout a system: instruction for goals, management, teaching and evaluation. Strategies could be more closely organized to reinforce and support each other.

Confusions

One of the concerns that bothers me this morning is the evident confusion between "minimum competency testing" and "competency based education." They are not the same. "Ideal" competency based education requires a major systems overhaul. Its characteristics are best described in Bill Spady's article in the Educational Researcher of January 1977.

Briefly stated, "ideal" competency based education calls for six critical adjustments in school management:

- o Explicitly stated, goal-oriented outcomes, not role-oriented outcomes
- o Flexible time for learning and evaluation, not fixed seat-time
- o A variety of in-school and out-of-school instructional activities that are goal-related
- o Measurement for diagnostic and remediation purposes, not for threat and control of student behavior
- o Certification based on performance on goals, not on attendance or age
- o Program adaptability, not program rigidity.

Minimum competency testing in its present form is an extremely narrow response to this ideal.

There is also a confusion that minimum competency tests test performance or the application of basic skills. They may or may not, depending upon the test. An exercise in which a student is asked to demonstrate the ability to punctuate by selecting an appropriate multiple choice answer falls short of a real-life application of that ability. Multiple choice tests are good for testing what Spady calls

"capacities" but this is a far cry from evaluation of the application of strings of competencies in real-life contexts. It's time for the testing community to begin to think about tests that are more holistic--real applied performance tests. This will necessitate continuous evaluation, begun early in elementary school with appropriate adjustment of learning opportunities to ensure competency acquisition. Evaluations would be made under more realistic circumstances than those afforded by multiple choice tests. There are good examples of more sophisticated evaluation techniques available.

Resources

There are several kinds of resources that are very important. Do you know about CAEL, the Council for Advancement of Experiential Learning? It's been devoting three or four years to non-traditional assessment techniques. A series of booklets can be ordered from them, including Willingham's six principles of the assessment of experiential learning, which would be very useful when assessing competencies at the secondary school level.

Do you know about CAPT, our host here today, which publishes a newsletter? It has been a proponent of applied performance testing and will continue that kind of work. It needs your help and contributions to continue spreading the word about applied performance testing.

Do you know about the work of Christine McGuire in the medical field? She works in the area of medical education and simulation. She has good pointers on how you can think about applied performance testing and testing of competencies in modes other than multiple choice testing.

Education is sadly behind other fields in developing assessment techniques. Business and industry have been using what is called the "assessment center technique." There is much written on that (although not in the educational assessment literature), but the methods can be useful, nevertheless.

The military, of course, has been involved in applied performance testing for years. The Human Resources Research Organization (HumRRO) is a good source of information on where to look for some ideas about testing competencies through performance.

Also, the booklet, Minimal Competency Testing: A Report of Four Regional Conferences, can be ordered through the Education Commission of the States. It contains a very exciting article by Mitch Brickel, entitled "Seven Key Notes on Minimal Competency Testing." It's very important, as we explore the meaning of this new movement towards minimum competency testing. And to restate my point, minimum competency testing is not ideal competency based education.

References

Anderson, Barry D. The costs of legislated minimum competency requirements. Background paper prepared for the Regional Workshops on Minimum Competency Testing, sponsored by the Education Commission of the States and the National Institute of Education, 1977.

Spady, William G. Competency based education: A bandwagon in search of a definition. Educational Researcher, 1977, 6 (1), 9-14.

Dale Carlson
California Department of Education

I'd like to compliment CAPT for putting on this conference, to thank the speakers for their good presentations, and to commend and thank the audience for your patience and courteous attention. I heard last night that listening was a basic skill. If so, you people have really mastered that one.

I would encourage Professor Walker to push his scenario technique to the absolute limit. I think that's a good way to surface all the implications of this movement. Maybe we could even put it in a research framework and study the impacts as they occur. Of course, what's actually going to happen will be especially interesting, because as we think about proficiency testing we know it's not possible! We really can't do it--that is, test for minimum competency with any reasonable degree of satisfaction to the various parties--and yet it's going to be done. It's going to be done because that's what the people are demanding. Yet it can't be done satisfactorily because it rests on logical conflict and value conflicts--the very reason debate rages so strongly. Even those responsible for developing minimum competency testing programs feel such ambivalence. This ambivalence comes from frustration in trying to reconcile conflicting principles--conflicts ranging from, "How can one have standards that are simultaneously uniform and variable? How can one have standards without identifying failures--or sacrificing the credibility of the standard or the standard setting process?" Most of these problems have roots to more fundamental conflicts in American values--for example, the conflict between our commitment to the

realization of the full potential of each person, and on the other hand, the deep seated, albeit often latent respect for egalitarian ideals with a healthy distrust for variability in that very human potential.

Merle's comments as written up various places are most useful to anyone on the firing line. They might save many people from lawsuits. I wish he would speak just a little bit, if time permits, about the opposite case in which a student does get a diploma but does not have what people might think are minimum competencies, that is, the Peter Doe type of case. That would allow us to see where we are, not just on that particular litigation, but on the whole idea, as far as the courts are concerned. I think his concept of instructional validity versus curricular validity of the tests is definitely worth elaborating and discussing. We'll no doubt have refinements on those ideas as time goes on, as programs develop and as problems unfold.

Carol and Lorrie, of course, lived up to my expectations, which are far above the minimum! I worry about Carol putting quite that much emphasis on reliance on judgment (although she probably did it today just for the sake of debate). Since we live by it day by day, we don't want to denigrate human judgment, but I think it would perhaps be safer to think of it as a bridge to something more lasting, such as the development of guidelines for verifying and evaluating the results of the judgment process itself.

And Lorrie, what can we say about the Florida experience? One could take any point of view. I'm speaking of the aspect where students did so much better on reading than they did in math, thereby causing most of the money and attention to go toward additional math instruction. What's interesting, I think, is that nobody really knows if it was

mainly a matter of imbalance in difficulty of test items between reading and math or a "true" difference in performance. It looks as if the people in Florida said, "Well, they scored lower on math items and hence are in greater need of math remediation." The same reading-math discrepancy has appeared in California's data as well. Whether it's because people pick harder math skills, or because math teachers have higher expectations, or whatever else, kids don't do as well on math items. But what is the truth? The normative data is enlightening but it doesn't automatically lead one to the truth any more than absolute judgment does--setting a standard at the 65 percent or 75 percent level. If the standard is adjusted to permit an equal percent failing, the decision may actually be just as arbitrary, although it appears to be more rational and enlightened. This Florida example clearly points out the difficulty in trying to set standards in this world where neither absolute nor normative data automatically lead to an answer. My thoughts call for research on minimum standards and how they can be rooted in success/failure ratios in various post-decision experiences for different types of students, but more about that later.

California has been mentioned a couple of times, so I'd like to spend just a minute running through the three California programs so that the one or two people who may be interested in California will know what it's about. First, there is the California Assessment Program, which is a statewide testing program. In contrast to what Lorrie said about the assessment program in some other states, California's is indeed quite broad, although still basic skill oriented. We hope in addition to its other functions it can be useful in detecting the impact of basic skill minimum competency testing efforts within the state.

But it's still a testing program per se--with a group or program emphasis. Accordingly, it uses matrix sampling procedures to collect group level information.

Then secondly, there is the little test that Ruth referred to, the California High School Proficiency Examination. It is an early exit test which students can take as early as the second semester of their sophomore year. If they pass, they are awarded a certificate of proficiency which is by law equal to a high school diploma, with all the rights and privileges pertaining thereto. The standard setting process was very simple. The author of the legislation thought that an average senior should pass it, so we just normed it on California high school seniors. We don't hold that up as an exemplary method to use. I only say that so if you hear rumors to the contrary, you'll know how it was really done. We can talk about the exam itself or the impact of it later if you like. The National Institute of Education has given us a grant to study the exam's impact on students who have taken it and who either have or have not left school, the results of which are being reported at an AERA symposium this year.

The third part is the minimum standards testing that will be going on very shortly in California. Districts now are in the process of trying to determine how they're going to carry it out. It is indeed a district oriented program. The state, for the first time in my recollection, does not have a central authority role; merely one of advising and assisting by putting out technical assistance guides, working with districts, and developing and disseminating examples of assessment exercises. The initial legislation stated that students must be tested or examined at least once between grades 7 through 9 and

twice in grades 10 through 12 unless they pass the exam the first time, of course. So it started out to be a high school minimum competency testing program. With legislation this last year it has been moved down to include testing in grades 4 through 6, thus meeting the criticism against testing only at the final decision point. So the change makes it a little more likely that the testing will become an integral part of the school instructional programs.

The particular legislation that moves it down to grades 4 through 6 was part of a major reform program in the schools in California, an upward extension of an earlier program of early childhood education. Now it extends through the upper elementary, junior high and senior high levels. Districts are awarded funds to plan and implement programs that are oriented toward basic skills, with emphasis on individualized learning, parent advising, and community involvement. So those are three things about California that you may or may not have known before, or cared about.

I'd like to retreat just for a moment to the sanctity of psychometric considerations, which seems only appropriate at a research and testing-related conference anyway. My plea is for research on this whole matter of standards. Judgment is important, but I think it would be more valuable to direct that judgment toward designing research studies and gathering data about the whole standard setting process. This implies faith in standards--that there may be something more or less invariant about standards--that there really are such things--which is obviously not true in any universal sense, but within well defined contexts such standards may have meaning.

I would hope we're on the verge of all kinds of research on standard setting. All the research on judgment procedures in psychology, decision theory, and all the operations research techniques undoubtedly have a bearing on this business of standard setting. There was a good session yesterday over in the Sheraton--the first session I've ever seen where actual data on the standard setting process was reported. Incidentally, we're reporting on some work in this area at a session tomorrow. Now I know ETS and others have been collecting data and I'd certainly encourage you to share it. How many people have done any kind of research, even involving only two or three teachers and 20 kids? Any kind of research on the standards setting process per se? Hold up your hands just for the fun of it. Okay, well it's a good start. See, if we all knew what each of you found, we'd be that much further ahead.

Thus, I make my plea for research on this whole process of setting standards. One could start with the judgment process itself. To compare different judgmental methods, one could investigate the impact of having people set standards under various conditions--by varying the amount of information available, varying the types of students, by doing it alone vs. together--to name a few ways. The variations go on and on. I think it would be even more useful to compare those techniques, as we tried to do, with the criterion groups method to see what happens when you do it in different ways. I think the criterion groups method is one useful step in this process of uncovering some ultimate truth about standards.

Then I'd really like to get into that third area where one searches for "real" standards. Obviously, it's all a probabalistic matter. The higher kids score on a test, the more likely they are to do better in some subsequent activity--an additional training program, college,

on-the-job or making a change in life. The real question about standards is whether there is some kind of critical point. Is there a point where the curve changes rapidly, above or below which the probability of success in subsequent activities varies precipitously? I haven't seen any data on that. I think it would be an interesting finding which would give standards some kind of essence in and of themselves. Lorrie has a little quote on standards being choices, not essences. I guess I'm hoping that what with all the pain and trouble people are going to, there's a little bit of essence somewhere!

As an example of this search for realness of standards, one thinks of the cloze technique. Bormuth, and other people following in his footsteps--the Madison, Wisconsin study especially--showed that at about a 35 percent correct score on a cloze test you see a rapid rise in the amount of information gained from reading. This indicates that there is some point below which students really don't understand what they're reading and above which they're able to learn from what they read. Maybe there are some analogous examples of that in mathematics which haven't yet been well explored.

Then there's the matter of the skills themselves. Which ones are really necessary? People will be setting all kinds of standards around the country on a variety of different types of skills, based on pure judgment. Some of those may be virtually unrelated to success in life or in some other training program, which seems to be a shame. Some of them may be related to success but still are far from indispensable. So we get into this whole area of compensatable skills, where some skills that seem to be related to future success are not really essential;

students can compensate for their deficit with a good personality or by using some other cognitive short-cut.

Lorrie pointed out one last thing that I hadn't thought of in preparing my remarks, that is the matter of looking at skills in terms of the impact schools can have on imparting them as a criterion for choosing which skills ought to be tested. This thought deserves much further consideration.

These are my thoughts. I thank you very much.

PART THREE

QUESTIONS AND ANSWERS

Question from audience member: I have a question to Carol Dwyer and Merle McClung because I think it involves a technical legal decision. It seems that the ETS presenter was basing the setting of standards on a secure test model like the college board or the Graduate Record Examination, the law school admissions test, or certification tests for architects--any profession. Most big external secondary school certification examinations run in the British/European system or the New York State system, that is, the test is secure only until the day it's given. After that it becomes public. You can buy it at the stationery office and, in fact, it's the tradition of the exam that influences curriculum and instruction. Now, if the test cannot be secure except up to the point at which it's given, it seems that the standards setting problem is complicated because you've got to go through the same procedure every year with different panels and there's going to be the question of equating standards from one year to another because of panel differences, test differences and so on. If that's the case, that in fact it may not be able to be secure, what happens then to the standards setting problem?

Dwyer: Maybe I could respond just by saying that in the approach Basic Skills Assessment has taken to that problem, we have developed a series of very closely parallel and statistically equated tests which move from the secure to insecure status on a periodic rotation. I don't see any problems with setting a standard on one form of the test and applying it to administration of another form of the test because of the

equating and the conceptual parallelism between subsequent forms of the test, or actually among all of the tests. We may find ourselves in the position where, by law, secure tests are not allowed. In which case we're faced with a problem that I think is not stricly a standards setting problem but is a more profound measurement problem. That is, what are the implications of prior knowledge of test materials for interpretation of test results?

I think there's really a generalized ability issue here too and it's the same one that comes up in curricular validity. That is, at what level of specificity do these things have to be demonstrated? We published detailed specifications. Is that sufficient? Or is knowing the kinds of items that are in a particular form are constructed to be parallel to another form sufficient? Or does the student have a right both to be taught and to examine the actual items? If it comes down to the conclusion that the student has a right to be taught those particular items, then test results mean something very different than they do in the testing situation that we're used to.

Question from audience member: I thought your presentation was remarkably clear but there was one point I was not sure of. When you were talking about not being able to envision standards for group decision making, I wasn't sure whether you were talking about standards for the process by which groups make decisions or whether you were talking about group and individual cut off points, etc. Could you clarify?

Shepard: I was talking about group cut-offs. That is, I would not evaluate a program in a school, in a district, or in a state by comparing aggregate information from that school or district in relation to absolute

standards. That means, I would not decide that a group was doing well in math or poorly in math by comparing their average performance to a standard.

Question from audience member: I've got two questions. The first one concerns the problem of having a standard error of measurement in a test and yet determining an absolute cut off score. In relation to whether someone gets a diploma or not, for instance. That is, as I understand it, any decision would have to be couched in such terms as we are 99 percent sure or 98 percent sure that this person can or cannot meet the standard. There's still a 2 percent chance of being wrong, in other words. Do these same problems come up as far as you know in terms of setting standards in particular companies for students who have graduated from a school and are now seeking employment?

Shepard: Any time you want to identify the most able by a valid criterion, you don't have the standard problem. All you have to do is demonstrate that the test is relevant to the job and then rank order the candidates on that criterion. You just start from the top and take them until you run out of positions.

Question from audience member: This is to Carol. I have no quarrel with setting some standards, whatever that means. I guess my concern is what are the implications for the setting of standards to the improvement of the curriculum and the instruction as it affects the youngster directly? More important, or concomitantly, what are the implications for teacher training? If we, in a mobile society and a changing society can assume that there are some standards that are universal, then what is the evidence, for example, that will show that

establishing minimum standards in fact will increase and improve the performance of youngsters curricular-wise and instruction-wise?

Dwyer: If I'm interpreting your question correctly, it sounds as if you're asking the opposite of what I get asked all the time, which is "how do you stop the minimum from becoming a maximum?" I assume that what you're interested in is allowing individuals to rise to their own levels of capacity whether these are students learning or teachers teaching the best way they can.

Audience member: The total improvement is what I'm interested in. With increased class size and the load on teachers, with the mobility, with the people changing their jobs--I'm not concerned about establishing some standards. What I'm concerned about is how will that really improve the quality of instruction for those youngsters in that community?

Dwyer: I think if the community is taking a close look at the rationale for its minimal competency testing program instead of just being reactive either to a bandwagon approach or to legislative pressures, what we're seeing is an examination of the total teaching process. I think that's always to the benefit of education when people can step back from their daily classroom duties, their daily administrative duties, and take a look at the big picture, of which testing is only a part. I think you're probably referring to the danger of minimal competency being applied as a sort of mindless standard.

Audience member: The other part of the question is whether there is some evidence right now that shows that the establishment of minimum competencies does in fact improve the total performance?

Dwyer: I don't think so.

Question from audience member: I probably have something that's halfway between a comment and a question, but it seems to be relevant. As test developers I think we tend to put psychological blinders on ourselves when we address this question. We assume that we somehow have to give a test that makes a judgment about competency that's a categorical imperative, that stands for all time, and that the test has to be given at one point in time. It seems to me that this is the exact opposite of the kinds of successful performance tests that we now have. Who of us would like to ride on an airplane where the pilot had not passed the competency test prior to our riding? Those types of tests are characterized by being public. Everybody knows exactly what the test entails and is able to take the thing as many times, within practical limits, as one wants in order to achieve success. It seems to me that we have to address exactly this type of problem and I was particularly interested in whether ETS has attempted this kind of thing. I understand that there is some movement in the sequential testing area.

Dwyer: I didn't really want to dwell too much on Basic Skills Assessment per se and in my remarks I really concentrated, for the purpose of discussion, on setting standards. In fact, the avowed purpose developed from our consortium of school districts (who were instrumental in putting the program together) was to identify students who are in need of remediation in order to reach an acceptable, locally-set level of performance for graduation. The main mode in which we recommend testing is testing as early as possible in the junior high or secondary years and subsequent opportunities for retesting with phases of remediation in

between retesting. And that is certainly a good use of the program in its secure form. What we've also done is released one of the earliest forms of the secure test that school districts can now keep within the district. They can score it by hand and allow the students many intervals in between official testing to check progress on their skills. As I said, I was emphasizing standards to delineate a contrast for the debate, but we really are very interested in multiple testing opportunities for identifying students needing remediation.

Question from audience member: This is for Loretta and it's kind of a three-pronged inquiry [regarding] your insistence that we fall back on a reliance on normative data. First of all, if it's really true that Johnny and Joanie can't read, write or reckon as well as they used to be able to do even in Palo Alto, [and] even if we renormed standardized tests at a halfway reasonable rate, I fail to see how there can be anything but a gradual drifting--anything that we would want to accept as a meaningful standard. I have a lot more reliance on the ability of human judgment to at least maintain some kind of standard over time. Secondly, the other thing that troubles me about normative data is that it almost inherently, inevitably demands that we set a cutting score somewhere in the tail of the distribution. I would think a realistic minimum competency model in a reasonably good setting, with dedicated teachers, a decent curriculum, motivated students, and not too many severe retardates admitted to the group should allow for the fact that everybody passes. That's what I would think minimum competency would mean in a good system. The third thing that troubles me is that those standardized tests that give normative data fall back on a nice normal curve. I would think that the kind of score distribution we'd want for

a minimal competency testing program would be a lovely negative skew where you minimize your false negatives and positives by stretching it out there at the lower end.

Shepard: I have several points to make. First of all, I am skeptical about human judgment because the reason we are in this fix is because of grade inflation and judgments being made by public school teachers every day. They have decided that other values like self concept and rewarding students for making an effort rather than doing a certain amount of work, or reaching a certain standard if you will, are more important. Therefore they have altered the standards in the last 10 years so that grades don't mean what they used to mean. So I don't trust human judgment as necessarily more insightful than normative information. I am by no means arguing for the content presently reflected in standardized tests, nor am I arguing for normal distributions when I suggest we look at present performance as one of the ways of informing such judgments. I'm not into that at all, I'm not arguing for that. I applaud the more careful specification of content in what we now call criterion referenced or, better still, domain referenced tests, and I expect a negative skew as well. Those two normal curves that I drew were those that are presented by mastery testing people who believe that mastery is an all or nothing phenomenon and that the normal distributions are errors of measurement distributed around either the mastery point or the non-mastery point.

Question from audience member: I'd like to respond to something that Merle said earlier on softening the blow of denying a diploma. I'm from a school district that mandated basic competencies in 1974 to go

into effect with the graduating class of 1977. It is a local requirement, not a state requirement, and the only way we've found to soften the blow of denying the diploma has been to allow the students to participate in graduation exercises. This seems to have been a big deal with the kids to be able to put on that cap and gown and march even though they receive a blank instead of a diploma. There was a precedent set for this because we allow students who have financial obligations that they have not met to pass, such as if they lost an algebra book or history book and they haven't paid for it yet. You can march but you won't get your diploma until you've paid for that. So, we've allowed the students to march but they have to come back to summer school or adult evening school and meet the requirement. Whenever they meet the requirement we then give them the diploma. So, what I'm saying is, Merle, we haven't found a way to soften the blow. Either you get it or you don't, depending on whether you pass the test.

Dwyer: Could I just make a comment? What you're describing is not really an all or nothing situation. It's all or nothing at one point, but you are in effect providing for upgrading of credentials because you're giving people a chance to come back and get their diploma.

Audience member: Right, and as you know we have remediation all along the way. Second grade checkpoint, fifth grade checkpoint, seventh grade checkpoint, and the proficiency exams three times a year at high school. Any time you fail you are placed in a remedial class and you have the opportunity to take the exam again. We have really attempted to plug all the loopholes, although I think there may be some out there that we didn't quite plug up. EMR students and LD students are not required to

pass the exams, and the exam is now in braille and in Spanish. Letters to parents come off the computer along with test results so that automatically parents are notified if a student passed or the student did not. So we have really attempted to plug up all the legal loopholes but we're sure there's someone out there who's just waiting.

Question from audience member: I'm Tom Knapp from Rochester. I have a question primarily for Lorrie Shepard and Merle McClung. I can't understand why people are so uptight about the human beings making cut off decisions. As I understand it, the person to be President of the United States must be 35 years old and a citizen. As far as I know, no one has sued for discrimination in that they weren't a citizen or they were 33 years old. Ruth mentioned the Boy Scouts as an example. I think you have to be at least 11 years old to be able to tie a couple of knots and recite the oath. Now, we have a number of these things in society and why should schools be immune to making such judgments of cut offs?

Shepard: My response, Tom, is just the severity of the consequences. I have no quarrel with mastery testing and individualized instruction in the classroom where we make these decisions almost daily or weekly. There's immediate redress when we make a mistake in such classifications. I think my concerns in this particular instance have to do with the cost of implementing such a program for high school graduation and the consequences of an incorrect classification.

Question from audience member: One short question. And thanks for an excellent conference. I wanted to ask, are there any good studies that indicate the effect of a prior model, a study of a prior model of a test,

on the scores? I'm asking that because I've seen several districts that are requiring certain kinds of tests. Then they're passing out copies of a similar kind of model to practice on. Some teachers are rejecting those models and others aren't and I wondered how it was going?

Dwyer: I don't know what to say. What I know about practice effects really deals with successive administrations of the same test, and my feeling about those is that that has always been a minor effect. I think most people, when they use practice tests, are really directing their attention at test anxiety rather than test performance per se. My own feeling would be that that's where the value lies in practice tests.

Nickse: I just want to mention the New York State Regents which has been giving exams for years and years. You can walk into any book store and buy the last ten years' tests in almost any subject matter area. If you're clever and if you have a good teacher, you can review those tests for the principles that are being tested. If you learn to do ten types of chemistry problems then you perhaps can hit a good score on the chemistry Regents for that year. So I think that though there probably isn't any formal study done, people do teach to the Regents test in New York. I find that not offensive as long as the tests are good and general principles are learned then by the students. How long they retain them is another problem.

Audience member: I have one or two comments for the panel, but first of all it has been very interesting and you brought up some of the problems we faced in Florida, some of which hopefully, we met. However, Lorrie,

I do want to clarify that our students are not fleeing to Georgia and we are surprised that Gene Glass used Time and Newsweek for his sources on such a paper.

At any rate, I do want to respond to the difference you pointed out in the percent of students passing the mathematics portion and the communication portion. I think you need to look at it in two ways. The communication test is a very easy test in terms of reading--reading such things as newspaper advertisements, and so forth. The mathematics test deals with solving everyday problems and it includes a great deal of reading. It also includes the basic skills in computation. Thus, it includes a student being able to read, being able to compute and being able to put it together. All the math curriculum people in the state are telling us that is what has happened on the math.

There's another thing that's happened on the math. Students read from eighth grade on continually. Math tests are math tests and the Functional Literacy Test measures basically 7th and 8th grade skills: percents, discounts and so forth. Some of the students who are now taking trigonometry and are college bound failed it because they had not utilized these skills for a number of years. With a review, of course, they will pass.

Another point I want to mention is that we do look for the percent of students failing Functional Literacy to drop after the next testing. They will have two more testings. The percent of failure should drop to about 4 or 5 percent of the students, based on some of the Denver Public Schools' findings.

I hope that you all would look at the Florida Accountability Act and the assessment program in relation to an overhaul of the whole

system, because that's essentially what it is. When I talk to groups, I really talk about the assessment program as a pivot between development and minimum performance standards, which is one section completely separate from our testing or assessment section. The Compensatory Education Act has provided $10 million for this year. That $10 million is only a semester's worth from January to June. The legislature thought that's all they'd be able to move--that schools could not get geared up before January. They have pledged $26.5 million next year.

That brings me to Dr. Walker's comment which I think someone else also mentioned. When he brought up the impact of putting those kinds of funds on the low ability students and the low socio-economic students, I hate to see the rich get richer and the poor get poorer. Have we thought about the cost to us and to society in terms of the cost of crime, the cost of prison systems and welfare? We are dealing with students who might not even be able to read the title on a bus in order to go to a place of employment, or to seek employment. So we feel the minimum competencies are very important. As far as setting the standards is concerned, there was a great deal of involvement from people all over the state. Teachers, curriculum people, and lay people all participated, so there has been quite a bit of consensus as to these minimal standards.

I know I said a lot and I know everybody's in kind of a hurry, but I do want to point out that I have a flyer with me on a conference we're having May 4th and 5th in Orlando in response to the many requests for information concerning our program. Dr. Popham will be there, there will be a banquet, and so forth. I have the information about the hotel and the registration procedures if any of you are interested. Also, I will be here if you have some questions about the Florida program.

Audience member: I'm Dave Orr from the National Center for Education Statistics in Washington. I'd like to offer a comment or two from what might be a different perspective here. I'm all for research and I'm all for psychometric technique and scientific accuracy and so forth. Indeed, that's my background. But I doubt that we're going to be able to sell scaled certification to school boards and legislatures any time soon. And indeed I'm afraid that we're doing what measurement people have been doing for as long as I can remember, and unfortunately that's getting longer than I like to admit. That is, we're building layers of statistical persiflage to hide hard decisions. If you think about this, minimum competency testing and standard setting and those related kinds of issues are not technical issues as much as they are socio-political issues. We are in this now because of socio-political pressures, not because somebody all of a sudden decided, "Hey, it would be really nice if we did some research on setting cutting scores for minimum standards." This is a socio-political problem and what I would like to hear people saying is, "Let's think of some socio-political ways to address the problem."

Now I'm not gainsaying the value of doing the statistical research, applying accepted principals of measurement and scientific method to the problem, but that's only part of it. Without addressing the socio-political issues I don't think it's going to be successful. Diplomas exist in a dichotomous fashion today. They are either awarded or they're not awarded. Operationally, that's a dichotomous variable. Sure, competencies underlie that and that may be a continuous variable but the kid has to deal with either I got it or I didn't. School boards and legislatures tend to look at it that way. As long as our measurement

procedures have classification error associated with them (and I don't think things have changed from back when they used to tell me in grad school a measurement error is axiomatic) and we depend on a cutting score or a psychometric procedure to award or not award, to certify or not certify, we're going to find ourselves in the next couple of years in the courts ad infinitum, ad nauseam. We're going to be spending our resources on defending legal cases instead of educating kids.

We have to look for a socio-political answer within which we can frame and couch the psychometric procedure. One possibility--I throw this out with not a great deal of thinking behind it--is a board of minimum competence at the district level, analagous to the typical college admission procedure. Sure, the SAT's form an important part of that. Sure, the minimum competency testing would form an important part of that. But the decision authority is vested in a board, just as the decision authority in college admissions is vested in an admissions office. You no longer are in the very bad position of having to go to court to defend a cutting score. So we need to think about socio-political approaches which will provide a buffer between ourselves and the legal system and still accomplish the job that we want to get done.